约束条件下 BN 参数最大熵模型扩展学习算法 *

郭文强†，李　然，侯勇严，高文强

(陕西科技大学 电气与信息工程学院，西安 710021)

摘　要：在很多智能系统的参数建模时，用户往往面对建模样本稀少的困境。针对在小数据集条件下贝叶斯网络（BN）参数建模的问题，提出了一种约束数据最大熵 BN 参数学习算法（CDME）。首先利用小数据集估算 BN 参数，随后把定性的专家经验转换为不等式约束，并利用 Bootstrap 算法生成满足约束的一组参数候选集，再根据信息最大熵进行加权计算出 BN 参数。实验结果表明，当数据量充分时，CDME 参数学习算法与经典的 MLE 算法的学习精度近似，表明了算法的正确性；在小数据集条件下，利用 CDME 算法，可以对 BN 进行参数建模，学习精度优于 MLE 算法和 QMAP 算法。CDME 算法在实际故障诊断样本数据相对稀缺的条件下，获取了诊断 BN 模型参数，在此基础上完成的诊断推理结果也印证了算法的有效性，为小数据集条件下的参数建模提供了一条新途径。

关键词：贝叶斯网络；小数据集；参数学习；最大熵模型

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2017.08.0868

Extension parameters learning for BN based on constraints and maximum entropy model

Guo Wenqiang†, Li Ran, Hou Yongyan, Gao Wenqiang

(*School of Electrical & Information Engineering Shaanxi University of Science & Technology*, *Xi'an* 710021 *China*)

Abstract: While the intelligent systems need parameter modeling, users often face the dilemma of scarce modeling samples. In this paper, a BN parameter learning method-constrained data maximum entropy (CDME) algorithm is proposed for the modeling of BN parameters under the small data sets. In the case of estimating BN parameters by using small data sets, the qualitative expert knowledge is transformed into inequality constraints for the sake of generating candidate parameter sets by Bootstrap algorithm. Then the BN parameters are estimated in the light of the maximum entropy principle. The experimental results show that CDME algorithm learning effects are similar to the classical MLE algorithm when the modeling data size is sufficient. However, when the data size is limited, the parameters of BN can be modeled by using the CDME, and the learnt accuracy is superior to MLE or QMAP algorithm. CDME is also applied to a real fault diagnosis while the data set is relatively scarce. The results of the diagnosis reasoning demonstrate that the presented parameter learning approach is effective. The CDME parameter learning algorithm provides a new modeling way for BN parameter under the small data sets.

Key Words: Bayesian Network; small dataset; parameter learning; maximum entropy model

0　引言

贝叶斯网络（Bayesian network，BN）在解决复杂系统的不确定性和不完整性问题凸显出强大的适应性，并且成功应用于智能系统涉及的众多领域[1~3]。BN 建模是 BN 推理的基础，而 BN 推理的结果也往往能检验出建模方法的有效性。BN 的建模包括结构建模和参数建模[4~5]。人们常常可以根据领域专家的经验获取一些定性知识，从而完成 BN 结构建模。而对于 BN 的参数建模，即估算 BN 的参数问题，专家往往很难直接确定出正确的定量参数。在样本数据充足的情况下，经典的最大似然估计（Maximum Likelihood Estimation, MLE）算法，是目前基于数据较为有效的学习方法之一。但在实际中，从某些系统中获取大量的有效样本数据非常困难和昂贵，如地震预测、航空发动机故障诊断系统中的样本数据等，这样就导致所能获得的系统特征数据相对较少，不能准确地估算出 BN 中的参数，对系统模型的推理产生影响，甚至得出与客观规律相违背的推理结果[6~8]。

在小数据集条件下 BN 的参数建模研究中，文献[9]利用定性约束提出了梯度下降估计（gradient-descent estimation, GDE）方法，利用 APN（adaptive probabilistic networks）算法进行求

基金项目：陕西省科技厅自然科学基金资助项目（2017JM6057，2013JK1114）

作者简介：郭文强（1971-），男（通信作者），陕西咸阳人，副教授，博士，主要研究方向为智能系统、机器学习等（guowenqiang@sust.edu.cn）；李然（1994-），男，硕士，主要研究方向为智能系统、模式识别等；侯勇严（1972-），女，陕西咸阳人，副教授，博士研究生，主要研究方向为智能控制等；高文强（1992-），男，硕士，主要研究方向为人工智能、模式识别等.

解。文献[10]针对定性影响约束提出了保序回归估计（isotonic regression estimation, IRE）方法，利用保序回归算法校正初始参数使得参数满足序约束。文献[11]针对规范性约束提出了最小自由能(Minimum Free Energy, MFE)估计方法，该方法首先将包含 KL 距离（Kullback–Leibler, divergence）和熵函数的最小自由能函数作为目标优化函数；然后，通过拉格朗日乘子将最小自由能函数与规范性约束结合起来，通过求极值的方法求取参数。文献[2]提出了利用定性影响约束进行 BN 参数学习的约束最大期望（constrained expectation maximum, CEM）方法。GDE、IRE、MFE、CEM 方法的约束类型单一，约束对参数的约束力较小，只能将参数限制到较为宽泛的可行域内，导致参数学习结果不够精确。文献[8]提出了基于双重约束的参数估计（dual constrained estimation, DCE）方法，利用 Beta 分布拟合和保序回归的方法进行小数据集条件下参数的双重约束学习。该算法适用于 BN 模型节点均为二值的情况。而现实系统中，如故障诊断等多模式分类问题，其网络节点状态往往是多值的，这就限制了 DCE 算法的应用。文献[12,13]提出了约束最大似然（constrained maximum likelihood, CML）方法。该方法首先构建非单调性约束模型；然后，将最大熵函数与非单调性约束结合构造凸优化问题然后求解。文献[14]提出了定性最大后验概率（qualitative maximum a posterior, QMAP）方法，然后，利用贝叶斯估计方法将真实统计数据和虚拟小数据相结合学习得到 BN 参数。

分析可以发现：CML 方法对参数约束类型没有特殊要求，对参数约束利用率较低，学习效果往往比 QMAP 逊色[12]。QMAP 方法是目前在小数据集条件下一种较为行之有效的新方法[8]。QMAP 方法利用定性约束通过拒绝—接受算法对满足约束的参数集合进行虚拟采样，而拒绝—接受算法是建立在概率分布函数估计的基础上，在实际中 BN 参数的概率分布函数往往很难估计，因而 QMAP 方法的实用性也受到挑战。

为了解决小数据集条件下 BN 参数的估计问题，本文提出了一种约束数据最大熵（constrained data maximum entropy, CDME）的参数学习算法。它把小数据集参数学习结果与利用专家经验约束扩展生成的参数候选集有机结合起来，在信息最大熵的基础上估计出 BN 的参数。

1 BN 参数学习背景知识

贝叶斯网络参数学习是在结构已知的情况下，通过样本数据和先验知识，获得所有网络节点的条件概率分布的过程。BN 由有向无环图(DAG)G和关联每个变量的概率分布集合组成。式（1）为 BN 的联合概率分布[1]。

$$P(X_1,X_2\cdots X_n)=\prod_{i=1}^{n}P(X_i\,|\,P_a(X_i)) \tag{1}$$

其中：$P_a(X_i)$ 表示 G 中 X_i 的父节点集合条件概率分布，$P(X_i\,|\,P_a(X_i))$ 表示包含在 G 中给定父节点值的变量的每个值的概率。

$P(X_i\,|\,P_a(X_i))$ 的第 k 个概率值表示为 $\theta_{ijk}=P(X_i=k\,|\,P_a(X_i)=j)$ 即是节点 X_i 的一个参数，其中，$\theta_{ijk}\in\theta$，$1\le i\le n$，$1\le j\le q_i$，$1\le k\le r_i$。r_i 表示 X_i 的状态数量，q_i 表示 X_i 父节点组合 $P_a(X_i)$ 的状态数量。显然，节点 X_i 一共有 $r_i\times q_i$ 个参数，它们构成了一个 $r_i\times q_i$ 维的矩阵，称之为节点 X_i 的条件概率分布表（condition probability table, CPT）。

1.1 最大似然估计算法

最大似然估计（MLE）表示为

$$L_\theta(e_1,e_2,\cdots e_m)=\prod_{i=1}^{m}f_\theta(e_i) \tag{2}$$

其中：$(e_1,e_2,\cdots e_m)$ 是样本集，θ 表示该分布的参数，$f_\theta(e_i)$ 为其概率密度函数[7]。特别是在实际中，可以用式（3）估计每个参数：

$$\theta_{ijk}^{MLE}=\frac{N_{ijk}}{N_{ij}} \tag{3}$$

其中：$P_a(X_i)$ 表示在样本 D 中，父节点 $P_a(X_i)$ 取第 j 个组合状态的频数，即 $N_{ij}=\sum_{k=1}^{r_i}N_{ijk}$。

MLE 常用于较充分数据集下 BN 参数估算，但式（3）有一个主要缺陷是不能估计 $N_{ij}=0$ 时的 θ_{ijk}^{MLE}。而当训练集有限时，观察值为 0 的现象时常发生，尤其是数据集规模较小时。为了解决这个问题，这里对式（3）做了一点修正如式（4）所示。

$$\theta_{ijk}^{*}=\frac{N_{ijk}+\varepsilon}{N_{ij}+u\varepsilon} \tag{4}$$

其中：$u=r_i\times q_i$；ε 为一个接近于零但不为零的常数。为满足实验要求，式（4）中 ε 取 0.00001。

1.2 常见的约束集合

在现实世界中，领域专家虽然很难给出 BN 模型参数的定量表达,但却往往可以提供节点间的定性信息，即约束信息。式（5）反映了专家先验知识（Ω 和 G）、数据（D）与 BN 参数之间的联系[11,12]。

$$\log P(\theta\,|\,G,D,\Omega)=\log(D\,|\,\theta,\Omega)+\log(G\,|\,\theta,\Omega)-C \tag{5}$$

其中：θ 是学习得到的 BN 参数，G 是 BN 结构，D 是样本数据，Ω 是通过专家经验知识形成的约束信息。

在 BN 中常见的约束集合可由定性领域知识转换而来。定性领域知识可以定义为条件概率上的不同约束，这些约束可以描述成以下约束形式[13,15]：

a）范围约束（range constraints）定义了 BN 中任意参数的上下界，$0<\alpha_{ijk}\le\theta_{ijk}\le\beta_{ijk}\le 1$。

b）交叉分布约束（cross-distribution constraints）定义了一对不同条件参数的相对关系。如果两个约束参数节点指数 i 和状态值 k，有不同的父节点组合 j，这个约束就被称为交叉分布约束，通常由定性因果关系得到，$\theta_{ijk}\le,\ge\theta_{ij'k},\forall j\ne j'$。

c）内分布约束（intra-distribution constraints）定义了一对相同条件的参数的相对关系。如果约束参数共享节点指数 i 和

父节点组合 j，却有不同的状态值 k，则这个约束被称为内分布约束，$\theta_{ijk} \leq,\geq \theta_{ijk'}, \forall k \neq k'$ 。

d）间分布约束（inter-distribution constraints）定义了一对相同条件的参数的相对关系。如果两个约束参数不共享任何节点指数 i、父节点组合 j 和状态值 k，这个约束被称为间分布约束，$\theta_{ijk} \leq,\geq \theta_{i'j'k'}, \forall i \neq i', j \neq j', k \neq k'$ 。

e）公理约束。描述参考固定父节点配置状态的参数之间的关系。这意味着，这是一个领域专家不需要提供且非常基本的约束，$\sum_{k=1}^{r_i} \theta_{ijk} = 1, 0 \leq \theta_{ijk} \leq 1, \forall i,j,k$ 。

f）近似等式约束。描述任何两个参数之间的密切关系。为了便于计算，因此可以将其写为：

$\left|\theta_{ijk} - \theta_{i'j'k'}\right| \leq \sigma, \forall i \neq i', j \neq j', k \neq k'$ 。

g）添加协同约束。描述不同分布下两个参数之和的比较关系：$\theta_{ij_1k} + \theta_{ij_2k} \leq \theta_{ij_3k} + \theta_{ij_4k}, \forall i,k$ 。

h）乘积协同约束。描述不同分布下两个参数的乘积之间的比较关系：$\theta_{ij_1k} * \theta_{ij_2k} \leq \theta_{ij_3k} * \theta_{ij_4k}, \forall i,k$ 。

1.3 最大熵原理

最大熵的基本思想是在满足所有已知事件的基础上对未知事件作最客观均一的估计。要从一个允许的概率分布集合 C 中选择一个模型，选择具有最大熵 $H(p)$ 的模型 $p_* \in C$，其模型的形式化表示为[16]

$$\begin{aligned} p_* &= \arg\max_{p \in C} H(p) \\ &= \arg\max_{p \in C} _ \sum_{x,y} p(x)p(y|x)\log p(y|x) \end{aligned} \tag{6}$$

其中：C 为满足约束条件的概率分布集合，x、y 为随机变量对应事件，条件分布 $p(y|x)$ 的数学测量由条件精提供。

本文选择最大熵统计模型作为原始领域（BN 知识和样本数据）和目标领域（BN 参数学习）的估计模型，主要基于以下考虑：最大熵所采用的 BN 分布参数对数据的先验分布也不作任何要求，可任意组合，而不会破坏整个估计模型的一致性；同时，它自然地解决了统计模型中参数平滑的问题，即 BN 参数应满足约束知识 Ω，假设有 L 条参数候选集都满足约束知识，根据最大熵原理，它们有均等的概率逼近真实的 BN 参数。因此，即使目标领域中部分特征未在原始领域出现，由于约束知识的支持，也能获得较好的参数估算效果

1.4 KL 距离

KL 距离，常用来衡量两个概率分布的距离。本文用 KL 距离来衡量 BN 参数估算方法的精度。如式（7）所示[17]。

$$KL(\theta_i | \theta_{i'}) = \sum_{j=1}^{q_i}\sum_{k=1}^{r_i} \theta_{ijk} \ln \frac{\theta_{ijk}}{\theta'_{ijk}} \tag{7}$$

其中：θ_{ijk} 代表真实参数，θ'_{ijk} 代表估算得到的参数。显然，估计的 BN 参数越接近真实 BN 参数，KL 值越小。

2 约束数据最大熵参数学习算法

本文提出的约束数据最大熵 BN 参数学习算法的流程如图 1 所示。

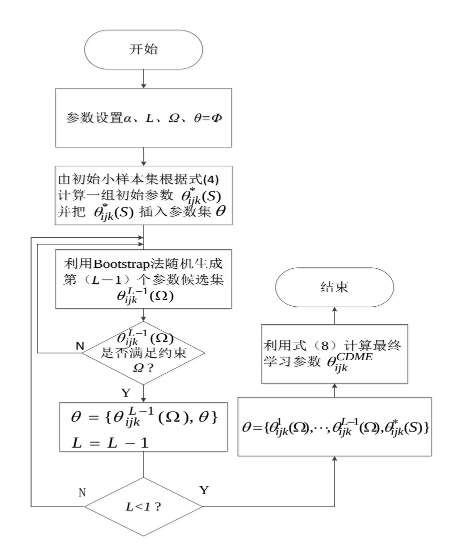


图 1　CDME 算法流程图

CDME 算法主要步骤如下：

a)设置参数权重因子 α（$0 \leq \alpha \leq 1$）；备选参数集个数 L、约束集合 Ω，令中间变量参数集 θ 为空集；

b)由初始小样本集根据式（4）计算出一组初始参数 $\theta^*_{ijk}(S)$，并把 $\theta^*_{ijk}(S)$ 插入参数集 θ；

c)利用 Bootstrap[18]方法生成第（L-1）个参数候选集 $\theta^{L-1}_{ijk}(\Omega)$ 。

d)判断第（L-1）个参数候选集 $\theta^{L-1}_{ijk}(\Omega)$ 是否满足约束 Ω。如满足，则把第（L-1）个参数候选集 $\theta^{L-1}_{ijk}(\Omega)$ 插入中间变量参数集 θ，并令 $L=L-1$；否则，返回第三步。

e)判断 L 是否小于 1。

（a）若是，输出，并将备选参数集 θ 参数取出，根据最大熵原则，按式（8）进行计算得到最终的学习参数 θ^{CDME}_{ijk} 。

$$\theta^{CDME}_{ijk} = \frac{\alpha \sum_{B=1}^{L-1} \theta^{(B)}_{ijk} + (1-\alpha)\theta^*_{ijk}(S)}{L} \tag{8}$$

其中:参数 $\sum_{B=1}^{L-1} \theta^{(B)}_{ijk}$ 是通过参数扩展得到的（L —1）个满足约束的参数候选集的和，$\theta^*_{ijk}(S)$ 是由初始小样本集计算出的一组初始参数，α 是权重因子。

（b）否则，跳转至步 c)。

注意到式（8）的参数模型中，当 α=0，L=1 时，在样本数据量充足条件下计算 BN 参数时，上式的形式实际是 MLE 的结果。因此该参数模型是 MLE 算法的泛化形式；MLE 是该参数模型的一种特例。

根据定性约束生成的候选参数具有同样的权重，体现了最大熵原则。另外，当 α 取较小值时，样本数据在参数学习结果中的决定性作用会凸显，而定性约束的专家经验对学习结果影响会减少；反之，当 α 取较大值时，定性约束的专家经验对学

习结果影响会增大，可补偿样本数据量少对学习结果的影响。

约束数据最大熵 BN 参数学习算法拟定了一个用约束信息和小数据集估算 BN 参数的框架。它可以将样本数据和专家定性经验等不同源的信息集中到一个框架下进行综合考虑，在解决 BN 参数问题时具有良好的形式上的一致性。

3　实验与分析

本文进行了两组实验分析来验证本文方法的有效性。第一组实验引用经典贝叶斯网络模型草坪湿润模型[19]在模拟数据下分别采用 MLE 方法、QMAP 方法和本文方法进行 BN 参数学习对比研究；第二组实验是在真实故障诊断数据下采用本文方法进行 BN 参数学习及推理诊断，进一步验证本文所提 BN 参数学习方法的优越性及实用性。本文所有实验所用 PC 配置为：2GHz 处理器，6G 内存，程序通过 MATLAB2014a 编程完成。

3.1　模拟数据 BN 参数建模实验

草坪湿润模型是验证 BN 学习算法的经典实验模型之一。图 2 为该 BN 模型的结构图。

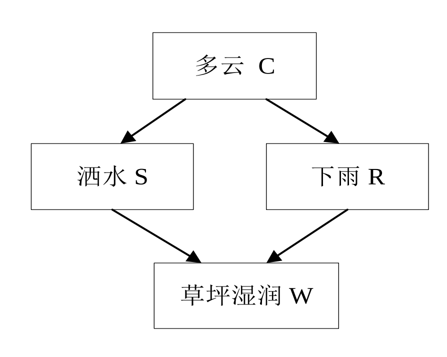


图 2　草坪湿润 BN 模型

图中该网络共有 4 个节点，分别是：C、S、R、W 节点，对 W 节点进行参数学习。

3.1.1 实验设置

参数权重因子 α 取 0.3，备选参数集个数 L 为 500，参数事件集合如表 1 所示，对节点 W 中的参数进行了编号。其中事件取值 “1” 表示事件为 “假”，“2” 表示事件为 “真”。表 2 为参数 P(W|S,R)所满足的约束集合 Ω。

表 1　参数事件集合

参数编号	事件
P1	P(W = 1 | S = 1, R = 1)
P2	P(W = 1 | S = 1, R = 2)
P3	P(W = 1 | S = 2, R = 1)
P4	P(W = 1 | S = 2, R = 2)
P5	P(W = 2 | S = 1, R = 1)
P6	P(W = 2 | S = 1, R = 2)
P7	P(W = 2 | S = 2, R = 1)
P8	P(W = 2 | S = 2, R = 2)

表 2　参数 P(W|S,R)所满足的约束集合

约束编号	约束条件	约束编号	约束条件
C1	P1>P5	C5	P7>P3,
C2	P8>P5	C6	P6>P5
C3	P8>P4	C7	P6>P2,
C4	P8>P7	C8	P7>P5,
C9	P8>P6	/	/

在表 2 中，P1>P5 表示在喷灌机（S）“没有喷水且天气（R）没有下雨”这一条件下，草坪事件 W 为“不湿”的概率大于“草地湿”的概率，本文将其称为约束条件“C1”，以此类推形成其他 8 个条件。很显然，上述约束集合是符合大众的认知，是能够被领域专家所欣然接受的。

3.1.2 实验结果分析

本文将在约束给定的情况下，给出不同样本数量下的 KL 距离对比，KL 距离通过式（7）计算得到。为了显示方便，本文对 KL 距离做了对数处理，形成相应的分贝值（dB），以此来衡量学习方法的学习精度。

在样本数量为 35 组的条件下，分别用 MLE 、QMAP、和 CDME 方法学习 W 节点参数。图 3 所示为三种方法学习 15 次得到的估计参数与真实参数间 KL 距离取平均值的对比情况。图 4 所示为相应的 KL 距离的盒形图。

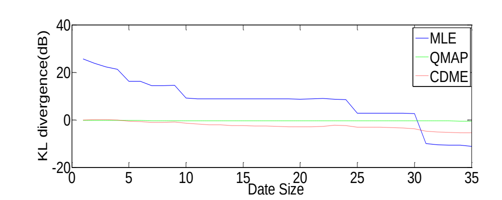


图 3　1~35 组样本数量下 W 节点三种方法 KL 距离对比

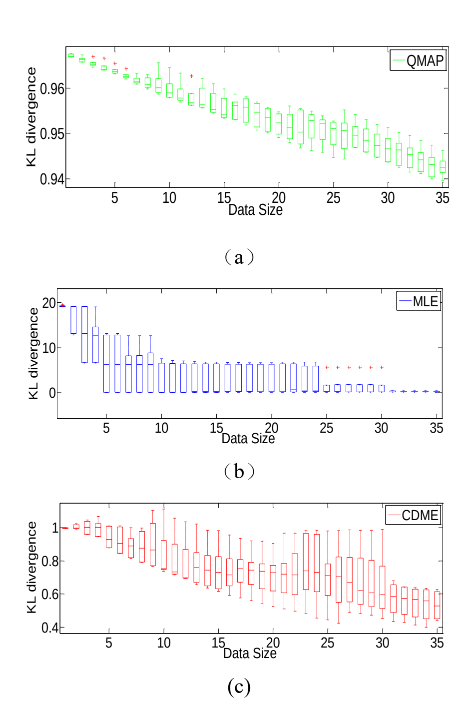


图 4　1~35 组样本数量下不同学习算法 KL 距离盒形图对比

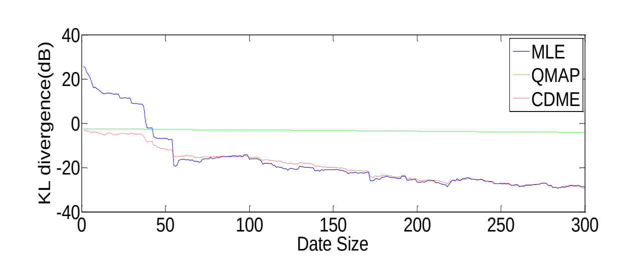


图 5　1~300 组样本数量下 W 节点三种方法 KL 距离对比

在样本数量相对充足时，实验中取 300 组样本，用三种方法学习 W 节点参数的 KL 距离对比如图 5 所示。

通过 BN 参数学习对比实验结果可以发现：在小数据样本数量下，MLE 方法 KL 距离较大，QMAP 方法和 CDME 方法具有明显优势，且 CDME 学习精度略高于 QMAP；随着样本数量的不断增加，MLE 方法 KL 距离值逐渐减小，并且在样本数量分别等于 30 和 31 时其 KL 距离与 QMAP 方法和 CDME 方法的 KL 距离相交。但 QMAP 方法的 KL 距离下降趋势不明显。与 QMAP 不同，CDME 方法随着样本数量增加，KL 距离总体呈一定的下降趋势。

分析可知，CDME 方法在小数据集条件下，可充分利用参数间的约束条件，估算出的 BN 参数在精度方面较 MLE、QMAP 方法更具优良的性能；在样本数量充足时本文方法依然可行，并能充分利用获取的数据集中蕴藏的信息，补偿 BN 建模的不确定性，不断修正、提高学习参数的精度。

3.2　真实数据下 BN 参数建模及故障诊断实验

本实验所用的轴承故障诊断模型和 469 条特征数据集均出自文献[20]。实验原始数据源自美国 Case Western Reserve University 轴承故障模拟实验室，他们用破坏性实验获取了丰富的典型故障类型的监测数据。而实际生产过程中，人们能获取的设备故障数据往往有限。为此，本文分别尝试在小数据（35 组）和充足数据（300 组）条件下，利用 CDME 算法进行 BN 参数学习，然后在此基础上，利用后 169 组特征数据集进行推理诊断实验，验证 CDME 算法的有效性。轴承故障诊断 BN 模型结构如图 6 所示。

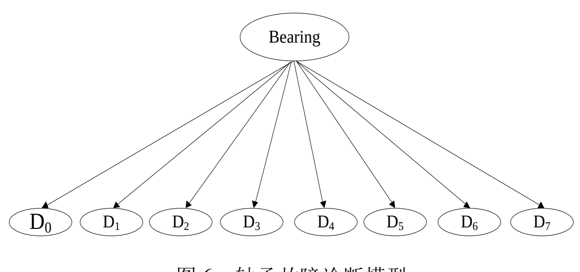


图 6　轴承故障诊断模型

图 6 中该 BN 模型共有 9 个节点，1 个父节点，8 个子节点代表 8 个特征向量 D_0~D_7。父节点有四种状态分别为：正常状态、内圈故障、外圈故障和滚动体故障。它们分别对应 BN 诊断模型的四种网络诊断输出，诊断信度为 γ=0.70。 为了对轴承故障进行诊断，先通过领域专家的经验完成了 BN 模型的结构建模。其次，对故障数据利用特征提取函数并离散化后得到特征向量，然后采取 CDME 方法进行故障诊断 BN 参数建模，得到故障诊断推理所需的 BN 模型。

3.2.1 实验参数设置

备用参数集个数 L 设为 500；参数约束集 Ω 根据专家经验得到 8 条；α 取 0.3。BN 推理选用经典的联合树算法。

3.2.2 故障诊断推理结果对比

表 3 和表 4 是分别利用 35 组、300 组特征数据建模,然后使用 169 组特征数据诊断推理结果。

表 3　35 组数据下 CDME 学习方法推理结果

状态类型	测试个数	正判个数	正判率
正常状态	169	169	100%
内圈故障	169	169	100%
滚动体故障	169	154	91.12%
外圈故障	169	139	82.25%

表 4　300 组数据下 CDME 学习方法推理结果

状态类型	测试个数	正判个数	正判率
正常状态	169	169	100%
内圈故障	169	169	100%
滚动体故障	169	162	95.86%
外圈故障	169	157	92.90%

实验结果表明：在小数据集条件下，用 CDME 算法建模之后进行推理，在整体上其正判率较高，推力能力较好（仅在推断外圈故障时其正判率略低于其他状态）。由此也进一步说明用 CDME 算法在小数据集条件下进行参数建模的有效性。在样本数量较充足时，CDME 算法可利用数据集信息修正 BN 学习参数，提高了 BN 模型的学习精度。

4　结束语

本文提出的 CDME 算法利用小数据集估算 BN 参数的同时，把定性的专家经验转换为不等式约束，利用 Bootstrap 算法生成满足约束的参数候选集，再根据信息最大熵原则计算出 BN 参数。实验表明：本文所提的 CDME 方法在小数据集下学习 BN 的参数具有一定的有效性和优越性。在故障诊断建模方面的应用，诊断推理结果印证了算法的实用性。在样本数量充足时本文算法依然可行，并能充分利用获取的数据集中蕴藏的信息，消减 BN 建模中的不确定性，不断修正、提高学习参数的精度。本文研究的 CDME 算法在实际智能系统建模数据稀缺的情况下，具有较好的实用前景，为小数据集条件下的 BN 参数建模提供了一种新途径。

参考文献：

[1] Dojer N. Learning Bayesian networks from datasets joining continuous and discrete variables [J]. International Journal of Approximate Reasoning, 2016,

78 (C): 116-124

[2] Liao Wenhui, Ji Qiang. Learning Bayesian network parameters under incomplete data with domain knowledge [J]. Pattern Recognition, 2009, 42 (11): 3046-3056.

[3] 陈为，朱标，张宏鑫. BN-Mapping: 基于贝叶斯网络的地理空间数据可视分析 [J]. 计算机学报, 2016, 39 (07): 1281-1293.

[4] 李硕豪，张军. 贝叶斯网络结构学习综述 [J]. 计算机应用研究, 2015, 32 (3): 641-646.

[5] 陈静，蒋正凯，付敬奇. 基于 Netica 的自学习贝叶斯网络的构建 [J]. 电子测量与仪器学报, 2016, 30 (11): 1687-1693.

[6] 肖蒙，张友鹏. 小数据集条件下的多态系统贝叶斯网络参数学习 [J]. 计算机科学, 2015, 42 (4): 253-257.

[7] 李子达，廖士中. 小样本贝叶斯网络参数学习方法 [J]. 计算机工程, 2016, 42 (8): 153-159+165.

[8] 郭志高，高晓光，邸若海. 小数据集条件下基于双重约束的 BN 参数学习 [J]. 自动化学报, 2014, 40 (7): 1509-1516.

[9] Altendorf E E, Restificar A C, Dietterich T G. Learning from Sparse Data by Exploiting Monotonicity Constraints [C]// Proc of the 21st Conference on Uncertainty in Artificial Intelligence. 2005: 18-25.

[10] Feelders A, Gaagl. Learning Bayesian networks parameters under order constraints [J]. International Journal of Approximate Reasoning, 2006, 42 (1//2): 37–53.

[11] Isozaki T, Kato N, Ueno M. “Data temperature”in minimum free energies for parameter learning of Bayesian networks [J]. International Journal on Artificial Intelligence Tools, 2009, 18 (5): 653–671.

[12] Niculescu R, Mitchell M, Rao B. Bayesian network learning with parameter constraints [J]. Journal of Machine Learning Research, 2006, 7 (1): 1357–1383.

[13] Campos C, Tong Yang, Ji Qiang. Constrained maximum likelihood learning of bayesian networks for facial action recognition [C]// Computer Vision. 2008: 168-181.

[14] Chang Rui, Shoemaker R, Wang Wei. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention [J]. Transactions on Computational Biology and Bioinformatics, 2011, 1 (8): 1170–1181.

[15] Guo Zhigao, Gao Xiaoguang, Di Ruohai, et al. Learning Bayesian network parameters from small data set: a spatially maximum a posteriori method [C]// Proc of International Workshop on Advanced Methodologies for Bayesian Networks. 2015: 32-45.

[16] Berger A L, Pietra S, Pietra V. A Maximum Entropy approach to Natural Language Processing [J]. Computational Linguistics, 1996, 22 (1): 38-73.

[17] Kullback S, Leibler R A. On information and sufficiency [J]. Annals of Mathematical Statistics, 1951, 22 (1): 79-86.

[18] 孙慧玲，胡伟文，刘海涛. 小样本情况下参数区间估计的改进方法 [J]. 哈尔滨理工大学学报, 2017, 22 (01): 109-113.

[19] Stuart J. Russell, Peter N. 人工智能：一种现代的方法 [M]. 殷建平，译. 3 版. 北京：清华大学出版社, 2013.

[20] 郭文强，张宝崃，彭程，等. 基于小波包和 BN 模型的深沟球轴承故障诊断 [J]. 轴承, 2016, 59 (3): 48-52.